

# Sparse Regression Codes for Multi-terminal Source and Channel Coding

Ramji Venkataramanan  
Dept. of Electrical Engineering  
Yale University, USA  
Email: ramji.venkataramanan@yale.edu

Sekhar Tatikonda  
Dept. of Electrical Engineering  
Yale University, USA  
Email: sekhar.tatikonda@yale.edu

**Abstract**—We study a new class of codes for Gaussian multi-terminal source and channel coding. These codes are designed using the statistical framework of high-dimensional linear regression and are called Sparse Superposition or Sparse Regression codes. Codewords are linear combinations of subsets of columns of a design matrix. These codes were introduced by Barron and Joseph and shown to achieve the channel capacity of AWGN channels with computationally feasible decoding. They have also recently been shown to achieve the optimal rate-distortion function for Gaussian sources. In this paper, we demonstrate how to implement random binning and superposition coding using sparse regression codes. In particular, with minimum-distance encoding/decoding it is shown that sparse regression codes attain the optimal information-theoretic limits for a variety of multi-terminal source and channel coding problems.

## I. INTRODUCTION

Among the important outstanding problems in network information theory is developing codes for various multi-terminal source and channel models that are provably rate-optimal with computationally efficient encoding and decoding algorithms. The introduction of deep ideas such as superposition [1], random binning [2] and auxiliary random variables [3]–[5] has led to a sharp characterization of information-theoretic limits for several network problems. However, until recently, even the best feasible codes for these problems fell short of these limits.

There have been some recent breakthroughs that begin to bridge this gap. Polar codes were the first codes with computationally feasible encoding algorithms that were shown to provably attain the information-theoretic limit for discrete-alphabet symmetric sources and channels [6]–[9]. Spatially coupled ensembles have recently been shown to achieve the capacity of binary-input symmetric-output channels with belief propagation decoding [10]. There are many important communication settings where the source or channel alphabet is inherently continuous, notably Gaussian sources and AWGN channels. Elegant techniques such as lattice coding have been proposed for continuous-alphabet source and channel coding [11]–[13], but these rate-optimal coding schemes do not have feasible encoding and decoding algorithms.

Recently a class of codes called Sparse Superposition Codes or Sparse Regression Codes (SPARC) was introduced by Barron and Joseph [14]–[16] for communication over the

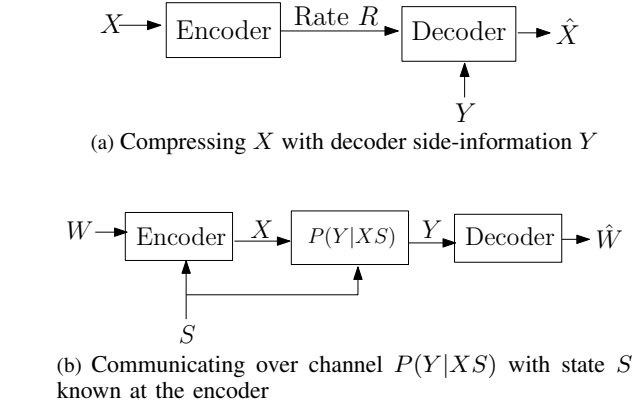


Fig. 1: Source and Channel Coding with Side-Information

AWGN channel. In [16], it was shown that SPARCs achieve the AWGN channel capacity with a computationally feasible decoding algorithm. SPARCs have also been shown to attain the optimal rate-distortion function of Gaussian sources with feasible algorithms [17]–[19]. In this paper, we show that the sparse regression framework can be used to design feasible codes for various Gaussian multi-terminal source and channel models.

The basic ingredients of the constructions used to prove coding theorems for many multi-terminal problems are:

- 1) Rate-optimal *point-to-point* source and channel codes,
- 2) Random binning,
- 3) Superposition coding.

As mentioned above, it has been shown in [14], [16], [18] that SPARCs are rate-optimal for Gaussian channels and sources. In this paper, we show that source and channel coding SPARCs can be combined to implement binning and superposition, thus yielding a new class of codes for multi-terminal source and channel coding. To illustrate how SPARCs can be used for binning, we consider the canonical examples of source coding with decoder side-information (the Wyner-Ziv problem [4]) and channel coding with encoder side-information (the Gelfand-Pinsker problem [5], [20]). These problems are depicted in Figure 1. Superposition coding using SPARCs is a natural extension of point-to-point coding and is illustrated via the Gaussian multiple-access and broadcast channels.

In Sections II and III, we review the SPARC construction and the minimum-distance performance results for source and channel coding. In Section IV, we describe how to implement random binning using SPARCs and use it to construct codes for the Wyner-Ziv problem (Figure 1a). The standard construction for this problem consists of a high-rate source codebook partitioned into bins, each of which serves as a lower-rate channel code. Due to the importance of the problem, several practical code-constructions have been proposed, e.g., [21]–[24], but they generally fall short of the Wyner-Ziv bounds; besides they do not come with provable performance guarantees. Recently polar codes have been proposed for Wyner-Ziv coding [7], [8]. These are the first computationally efficient code constructions that are provably rate-optimal. However, these are only applicable to problems where the source and side-information distributions are discrete and symmetric. Elegant coding schemes such as those based on lattices have been proposed [11], [25] for the Wyner-Ziv problem with continuous-valued source and side-information, but they have exponential encoding and decoding complexity.

In Section V, we turn our attention to channels with state, where the state information is known non-causally at the transmitter. This model (Figure 1b) has been studied widely in the literature [5], [11], [20], [26] and has found many practical applications such as multi-antenna communication [27], digital watermarking [28], [29] and steganography [30]. It is the channel coding dual of the Wyner-Ziv problem [31], [32]. In Figure 1b, the encoder knows the entire state sequence  $S$  at the beginning of communication while the decoder observes only the channel output  $Y$ . This capacity of this channel model was determined by Gelfand and Pinsker [5]. For the important special case of AWGN channels with Gaussian state, Costa [20] showed that the Gelfand-Pinsker capacity is the same as the rate achievable when the decoder has full knowledge of  $S$ . Since Costa's discovery of this surprising result (dubbed 'writing on dirty paper'), elegant capacity-achieving coding schemes have been developed such as nested lattice codes [11], [26], [29], but these are generally computationally infeasible. Several computationally efficient code designs have also been proposed, e.g., [33], [34]; however they do not come with provable rate guarantees. In Section V, we show how to implement Costa's coding scheme by partitioning a high-rate SPARC channel code into bins of lower-rate source codes.

Finally in Section VI, we show how to construct capacity-achieving codes for the AWGN multiple-access and broadcast channels using SPARCs. We show that superposition codes for these channels can be implemented through a simple extension of SPARCs for point-to-point channel coding.

The analysis of SPARCs in this paper is presented with minimum-distance encoding and decoding, which is optimal but computationally inefficient. This is mainly to keep exposition simple and to highlight the main contribution of the paper – a demonstration that binning and superposition can be easily implemented with sparse regression ensembles described by compact dictionaries. The results also hold with the feasible SPARC encoders and decoders developed in [16], [17], [19].

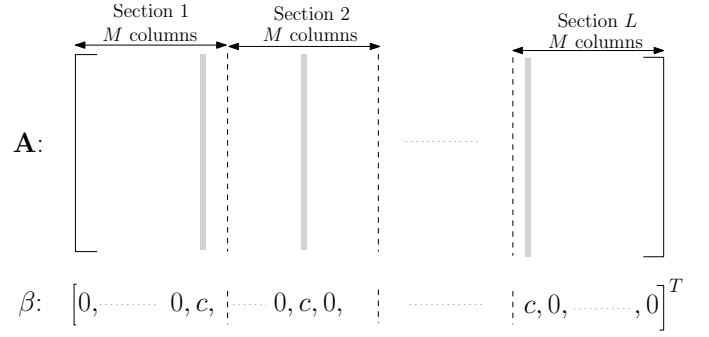


Fig. 2:  $\mathbf{A}$  is an  $n \times ML$  matrix and  $\beta$  is a  $ML \times 1$  binary vector. The positions of the non-zeros in  $\beta$  correspond to the gray columns of  $\mathbf{A}$  which add to form the codeword  $\mathbf{A}\beta$ .

Further, we focus only on the achievability of the optimal information-theoretic rates and do not discuss the SPARC error exponents obtained in [14], [18]. These aspects will be discussed in an extended version of this paper.

*Notation:* Upper-case letters are used to denote random variables, lower-case for their realizations, and bold-face letters to denote random vectors and matrices. All vectors have length  $n$ .  $\|\mathbf{X}\|$  denotes the  $\ell_2$ -norm of vector  $\mathbf{X}$ , and  $|\mathbf{X}| = \|\mathbf{X}\|/\sqrt{n}$  is the normalized version. We use natural logarithms, so entropy is measured in nats. To limit the number of symbols introduced, we reuse notation across sections. For example,  $X$  is used to represent the channel input as well as the source;  $Y$  is used to denote both the channel output and the source side-information. The model description at the beginning of each section explains all the variables used in it.

## II. SPARSE REGRESSION CODES

A sparse regression codebook (SPARC) is defined in terms of a design matrix  $\mathbf{A}$  of dimension  $n \times ML$  whose entries are i.i.d.  $\mathcal{N}(0, 1)$ , i.e., independent zero-mean Gaussian random variables with unit variance. Here  $n$  is the block length and  $M$  and  $L$  are integers whose values will be specified shortly in terms of  $n$  and the rate  $R$ . As shown in Figure 2, one can think of the matrix  $\mathbf{A}$  as composed of  $L$  sections with  $M$  columns each. Each codeword is a linear combination of  $L$  columns, with one column from each section. Formally, a codeword can be expressed as  $\mathbf{A}\beta$ , where  $\beta$  is an  $ML \times 1$  vector  $(\beta_1, \dots, \beta_{ML})$  with the following property: there is exactly one non-zero  $\beta_i$  for  $1 \leq i \leq M$ , one non-zero  $\beta_i$  for  $M + 1 \leq i \leq 2M$ , and so forth. Denote the set of all  $\beta$ 's that satisfy this property by  $\mathcal{B}_{M,L}$ . The non-zero values of  $\beta$  are all set equal to  $c \triangleq \frac{\gamma}{\sqrt{L}}$  where  $\gamma$  will be specified later depending on the problem at hand.

Since there are  $M$  columns in each of the  $L$  sections, the total number of codewords is  $M^L$ . To obtain a rate of  $R$  nats/sample, we therefore need

$$M^L = e^{nR}. \quad (1)$$

There are several choices for the pair  $(M, L)$  which satisfy this. For example,  $L = 1$  and  $M = e^{nR}$  recovers the Shannon-style random codebook; here the number of columns

in the dictionary  $\mathbf{A}$  is  $e^{nR}$ , i.e., exponential in  $n$ . For our constructions, we choose  $M = L^b$  for some  $b > 1$  so that (1) implies

$$L \log L = nR/b. \quad (2)$$

Thus  $L$  is now  $\Theta\left(\frac{n}{\log n}\right)$ , and the number of columns  $ML$  in the dictionary  $\mathbf{A}$  is now  $\Theta\left(\frac{n}{\log n}\right)^{b+1}$ , a polynomial in  $n$ . This reduction in dictionary complexity can be harnessed to develop computationally efficient encoders and decoders for the sparse regression code.

Since each codeword in a SPARC is a linear combination of  $L$  columns of  $\mathbf{A}$  (one from each section), codewords sharing one or more common columns in the sum will be dependent. Also, SPARCs are not linear codes since the sum of two codewords does not equal another codeword in general.

### III. SPARC FOR POINT-TO-POINT SOURCE AND CHANNEL CODING

In this section, we review the performance of SPARCs for point-to-point source and channel coding under minimum distance encoding/decoding.

#### A. Lossy Source Coding

Consider an i.i.d Gaussian source  $X$  with mean 0 and variance  $\sigma^2$ . A rate-distortion codebook with rate  $R$  and block length  $n$  is a set of  $e^{nR}$  length- $n$  codewords, denoted  $\{\hat{\mathbf{X}}(1), \dots, \hat{\mathbf{X}}(e^{nR})\}$ . The quality of reconstruction is measured through the mean-squared distortion criterion

$$d_n(\mathbf{X}, \hat{\mathbf{X}}) = \|\mathbf{X} - \hat{\mathbf{X}}\|^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)^2,$$

where  $\hat{\mathbf{X}}$  is the codeword chosen to represent the source sequence  $\mathbf{X}$ . For this distortion criterion, an optimal encoder maps each source sequence to the codeword nearest to it in Euclidean distance. The rate-distortion function  $R^*(D)$ , the minimum rate for which the distortion can be bounded by  $D$  with high-probability, is given by [35]

$$R^*(D) = \min_{p_{\hat{\mathbf{X}}|\mathbf{X}}: E(X - \hat{X})^2 \leq D} I(X; \hat{X}) = \frac{1}{2} \log \frac{\sigma^2}{D} \text{ nats/sample.} \quad (3)$$

For rates  $R > R^*(D)$ , a sparse regression codebook is defined in terms of an  $n \times ML$  design matrix  $\mathbf{A}\beta$ , as described in the previous section. The non-zero values of  $\beta \in \mathcal{B}_{M,L}$  are all set equal to  $\sqrt{(\sigma^2 - D)/L}$ . Encoding and decoding are as follows.

*Minimum-distance Encoder:* This is defined by a mapping  $g: \mathbb{R}^n \rightarrow \mathcal{B}_{M,L}$ . Given the source sequence  $\mathbf{X}$ , the encoder determines the  $\beta$  that produces the codeword closest in Euclidean distance, i.e.,

$$g(\mathbf{X}) = \underset{\beta \in \mathcal{B}_{M,L}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{A}\beta\|^2.$$

*Decoder:* This is a mapping  $h: \mathcal{B}_{M,L} \rightarrow \mathbb{R}^n$ . On receiving  $\beta \in \mathcal{B}_{M,L}$  from the encoder, the decoder produces reconstruction  $h(\beta) = \mathbf{A}\beta$ .

The probability of error at distortion-level  $D$  of a rate-distortion code  $\mathcal{C}_n$  with block length  $n$  and encoder and decoder mappings  $g, h$  is

$$P_e(\mathcal{C}_n, D) = P(|\mathbf{X} - h(g(\mathbf{X}))|^2 > D). \quad (4)$$

It was shown in [18] that SPARCs can achieve the optimal rate-distortion function with the optimal error-exponents for i.i.d Gaussian sources for all distortions  $D$  such that  $D/\sigma^2 < x^*$ , where  $x^* \approx 0.2032$  is the solution of the equation

$$1 + \frac{1}{2} \log x = x. \quad (5)$$

*Fact 1:* [18] For  $D \in (0, \sigma^2)$ , let  $R_{sp}(D) = \max\{\frac{1}{2} \log \frac{\sigma^2}{D}, 1 - \frac{D}{\sigma^2}\}$ . Fix rate  $R > R_{sp}(D)$ ,  $\epsilon > 0$  and  $b > b_1$  where

$$b_1 = \frac{2.5R}{R - 1 + D/\sigma^2}. \quad (6)$$

For all  $n$ , let  $\mathcal{C}_n$  be a rate  $R$  SPARC defined by an  $n \times L_n M_n$  design matrix with i.i.d  $\mathcal{N}(0, 1)$  entries, where  $L_n$  is determined by (2) and  $M_n = L_n^b$ . Then for all sufficiently large  $n$ ,  $P_e(\mathcal{C}_n, D) < \epsilon$ .

Fact 1 implies that SPARCs achieve the optimal rate-distortion function for  $0 < \frac{D}{\sigma^2} < x^*$  where  $x^* \approx 0.2032$  is the solution of (5). For  $x^* \leq \frac{D}{\sigma^2} \leq 1$ , the minimum achievable rate of Fact 1 ( $1 - \frac{D}{\sigma^2}$ ) is larger than the optimal rate-distortion function.

#### B. Communication over an AWGN Channel

Consider an AWGN channel with input  $X$  and output  $Y$  defined by

$$Y = X + Z$$

where  $Z \sim \mathcal{N}(0, N)$  is a noise variable independent of  $X$ . There is an average power constraint  $P$  on the input  $X$ . Denote by  $v$  the signal-to-noise ratio  $P/N$ . It was shown in [14] that SPARCs can achieve the capacity  $\frac{1}{2} \log(1 + v)$  with the probability of error decaying exponentially with  $n$ .

*Encoder:* This is a mapping  $g: \mathcal{B}_{M,L} \rightarrow \mathbb{R}^n$ . Each message in the set  $\{1, \dots, M^L = e^{nR}\}$  is indexed by a unique  $\beta \in \mathcal{B}_{M,L}$ . The non-zero values of  $\beta$  are all equal to  $\sqrt{P/L}$ . To transmit the message corresponding to  $\beta$ , the encoder produces the channel input  $\mathbf{X} = \mathbf{A}\beta$ .

*Minimum-distance Decoder:* This is defined by a mapping  $h: \mathbb{R}^n \rightarrow \mathcal{B}_{M,L}$ . Upon receiving the output sequence  $\mathbf{Y}$ , the decoder determines the  $\beta$  that produces the codeword closest in Euclidean distance, i.e.,

$$\hat{\beta} = h(\mathbf{Y}) = \underset{\beta \in \mathcal{B}_{M,L}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{A}\beta\|^2.$$

The average probability of error of a code  $\mathcal{C}_n$  with block length  $n$  and encoder and decoder mappings  $g, h$  is

$$P_e(\mathcal{C}_n) = \frac{1}{M^L} \sum_{\beta \in \mathcal{B}_{M,L}} P(\hat{\beta} \neq \beta | \mathbf{X} = \mathbf{A}\beta). \quad (7)$$

The performance of SPARC for channel coding is given below.

Let  $v^* \approx 15.8$  be the solution to  $(1+v^*)\log(1+v^*) = 3v^*$ . Define

$$b_0(v) = \begin{cases} \frac{4v(1+v)\log(1+v)}{[(1+v)\log(1+v)-v]^2} & v < v^* \\ \frac{(1+v)\log(1+v)}{(1+v)\log(1+v)-2v} & v \geq v^* \end{cases} \quad (8)$$

We note that  $b_0$  asymptotically approaches 1 with growing  $v$ .

*Fact 2:* [14] Fix rate  $R < C = \frac{1}{2} \log(1+v)$ ,  $b > b_0(v)$  and  $\epsilon > 0$ . For all  $n$ , let  $\mathcal{C}_n$  be a rate  $R$  SPARC defined by an  $n \times L_n M_n$  design matrix with i.i.d  $\mathcal{N}(0, 1)$  entries, where  $L_n$  is determined by (2) and  $M_n = L_n^b$ . Then for all sufficiently large  $n$ ,  $P_e(\mathcal{C}_n) < \epsilon$ .

#### IV. SPARC FOR LOSSY COMPRESSION WITH DECODER SIDE-INFORMATION

In this section, we construct SPARCs to achieve the optimal Wyner-Ziv rate for Gaussian sources. Consider an i.i.d Gaussian source  $X \sim \mathcal{N}(0, \sigma^2)$  to be compressed with mean-squared distortion  $D$ . The decoder side-information  $Y$  is noisy version of  $X$  and is related to  $X$  by  $Y = X + Z$ , where  $Z \sim \mathcal{N}(0, N)$  is independent of  $X$ . The sequence  $\mathbf{Y}$  is available at the decoder non-causally. If  $\mathbf{Y}$  were available at the encoder as well, the optimal strategy is to compress  $\mathbf{Z} = \mathbf{Y} - \mathbf{X}$  to within distortion  $D$ ; the minimum rate required for this is  $\frac{1}{2} \log \frac{\text{Var}(X|Y)}{D}$  nats/sample. Wyner and Ziv showed in [4] that this rate is achievable even when  $\mathbf{Y}$  is available at only the decoder.

Before presenting the SPARC construction, we briefly review the main ideas in the Wyner-Ziv random coding scheme [4]. Define an auxiliary random variable  $U$  jointly distributed with  $X$  according to

$$U = X + V \quad (9)$$

where  $V \sim \mathcal{N}(0, Q)$  is independent of  $X$ . The idea is that the decoder first recovers  $U$ , and then produces  $\hat{X}$  as the best estimate of  $X$  given  $U$  and  $Y$ . The codebook consists of length- $n$  vectors chosen i.i.d according to the marginal distribution of  $U$ . The encoder attempts to find a codeword  $\mathbf{U}$  whose empirical joint distribution with  $\mathbf{X}$  is close to (9). From the rate-distortion theorem, this step will be successful if the codebook size is at least slightly larger than  $e^{nI(U;X)}$ . Since the decoder has  $\mathbf{Y}$ , the index of the chosen codeword  $\mathbf{U}$  is not sent in its entirety; instead we divide the codebook into  $e^{nR}$  equal-sized bins and send only the index of the bin containing the codeword. Thus the information rate to the decoder is  $R$  nats/source sample which is less than the rate  $I(U; X)$  required to convey the precise codeword index.

The decoder's task is to recover the codeword  $\mathbf{U}$  using the bin index and the side-information  $\mathbf{Y}$ . This is equivalent to a *channel decoding* problem. We can correctly distinguish  $\mathbf{U}$  from the other codewords in the bin if number of codewords in each bin is exponentially less than  $e^{nI(U;Y)}$ . Combining this with the minimum codebook size for quantization, we see

that the number of bins  $e^{nR}$  should satisfy

$$e^{nR} > \frac{e^{nI(U;X)}}{e^{nI(U;Y)}}$$

or

$$R > I(U; X) - I(U; Y) = \frac{1}{2} \log \frac{\text{Var}(X|Y)}{D}$$

where the last inequality is obtained by setting  $Q = \frac{\text{Var}(X|Y)D}{\text{Var}(X|Y)-D}$ . After decoding  $\mathbf{U}$ , the decoder reconstructs  $\hat{\mathbf{X}}$  as the MMSE estimate of  $\mathbf{X}$  given  $(\mathbf{U}, \mathbf{Y})$ . It can be verified that expected squared-error distortion is  $D$ .

We now show that the above coding scheme with binning can be implemented with SPARCs. The relation (9) can be equivalently written in terms of the reverse test channel as

$$X = aU + V' \quad (10)$$

where  $a = \frac{\sigma^2}{\sigma^2+Q}$ , and  $V' \sim \mathcal{N}(0, \frac{\sigma^2 Q}{\sigma^2+Q})$  is independent of  $U$ . The first step of the coding scheme is equivalent to quantizing the source sequence  $\mathbf{X}$  to a codeword  $a\mathbf{U}$  with mean-squared distortion at most  $\frac{\sigma^2 Q}{\sigma^2+Q}$ . We can use a SPARC to perform this quantization by choosing a design matrix with parameters satisfying the specifications in Fact 1.

Instead of sending the codeword index  $\beta$  to the decoder in its entirety, we divide each section of the design matrix  $\mathbf{A}$  into subsections of  $M'$  columns each as shown in Figure 3, and only send information to indicate which subsection in each of the  $L$  sections of  $\beta$  contains a non-zero. More precisely, we send the decoder a tuple  $(p_1, \dots, p_L)$  where  $p_i \in \{1, \dots, \frac{M}{M'}\}$  indicates a subsection in the  $i$ th section of  $\mathbf{A}$ . This strategy is equivalent to binning: a bin is now a subset of the codebook consisting of codewords corresponding to  $\beta$ 's with ones in the sections specified by  $(p_1, \dots, p_L)$ . The codebook is thus divided into  $(\frac{M}{M'})^L$  bins and the rate  $R$  required to send the bin index to the decoder is determined as

$$e^{nR} = (M/M')^L. \quad (11)$$

We note that each bin is itself a smaller sparse superposition codebook with  $M'^L$  codewords, defined by a  $n \times M' L$  sub-matrix of  $\mathbf{A}$ .

The decoder side-information variable  $Y$  is related to  $U$  as

$$Y = X + Z = aU + V' + Z \quad (12)$$

where  $U, V'$  and  $Z$  are mutually independent. The problem of recovering  $U$  from  $Y$  at the decoder is a channel decoding problem over a channel with signal-to-noise ratio given by

$$\text{snr} = \frac{a^2 \text{Var}(U)}{\text{Var}(V') + \text{Var}(Z)} = \frac{\sigma^4}{\sigma^2 Q + (\sigma^2 + Q)N}. \quad (13)$$

Since Fact 2 shows shown that SPARCs can achieve the AWGN channel capacity, the decoder can perfectly recover  $\mathbf{U}$  if the number of codewords in each bin satisfies

$$M'^L < \exp\left(n \frac{1}{2} \log(1 + \text{snr})\right).$$

The above SPARC coding scheme and its performance are formalized below.

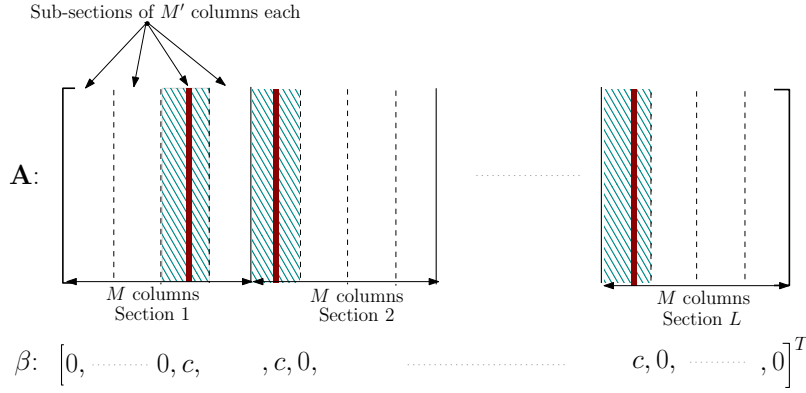


Fig. 3: Each section is divided into subsections of  $M'$  columns. A bin is formed by specifying a subsection in each of the  $L$  sections as shown by the shaded regions.

**Definition 1 (Nested Sparse Regression Codebook):** A nested sparse regression codebook with rates  $(R_1, R_2)$  and block length  $n$  is defined by an  $n \times ML$  design matrix  $\mathbf{A}$  with i.i.d  $\mathcal{N}(0, 1)$  entries, where  $M^L = e^{nR_1}$ . Each section of  $M$  columns is divided into sub-sections of  $M'$  columns each, where  $M'$  is determined by  $M'^L = e^{nR_2}$ . The codebook consists of codewords  $\mathbf{A}\beta$ , where  $\beta \in \mathcal{B}_{M,L}$  contains one non-zero element in each of the  $L$  sections.

**Theorem 1:** Fix  $R_1 > \max\{\frac{1}{2} \log \frac{\sigma^2 + Q}{Q}, \frac{\sigma^2}{\sigma^2 + Q}\}$  and  $R_2 < \frac{1}{2} \log(1 + \text{snr})$  where

$$Q = \frac{\text{Var}(X|Y)D}{\text{Var}(X|Y) - D}$$

and  $\text{snr}$  is given by (13). Then for any  $\epsilon > 0$  and all sufficiently large  $n$ , there exists a rate  $R_1 - R_2$  code  $\mathcal{C}_n$  with  $P_e(\mathcal{C}_n, D) < \epsilon$  where  $\mathcal{C}_n$  is defined by a nested sparse regression codebook with rates  $(R_1, R_2)$  whose  $n \times ML$  design matrix satisfies the following:  $M = L^b$  where

$$b > \max \left\{ \frac{2.5R_1}{R_1 - \sigma^2/(\sigma^2 + Q)}, \frac{R_1}{R_2} b_0(\text{snr}) \right\}$$

and  $L$  is determined by  $bL \log L = nR_1$ . (The function  $b_0(\cdot)$  is defined in (8)).

Define

$$D^* = \frac{x^* \sigma^2}{1 + x^* \sigma^2 / N}$$

where  $x^* \approx 0.2032$  is the solution of (5).

**Corollary 1:** For  $D \in (0, D^*)$ , sparse regression codes achieve the optimal Wyner-Ziv rate-distortion function for Gaussian sources given by  $\frac{1}{2} \log \frac{\text{Var}(X|Y)}{D}$ .

**Proof:** It can be verified that for  $D \in (0, D^*)$ , the lower bound on  $R_1$  specified by the theorem becomes  $\frac{1}{2} \log \frac{\sigma^2 + Q}{Q}$ . The corollary then follows by choosing  $R_1 = \frac{1}{2} \log \frac{\sigma^2 + Q}{Q} + \epsilon$  and  $R_2 = \frac{1}{2} \log(1 + \text{snr}) - \epsilon$ . This yields an achievable rate  $R_1 - R_2 = \frac{1}{2} \log \frac{\text{Var}(X|Y)}{D} + 2\epsilon$  where  $\epsilon > 0$  can be arbitrarily small. ■

**Proof of Theorem 1:**

Fix block length  $n$  and rates  $R_1, R_2$ . Choose a  $n \times ML$  design matrix  $\mathbf{A}$  with  $M = L^b$  and  $bL \log L = nR_1$  where  $b$  is

greater than the minimum value specified by the theorem. Each section of  $\mathbf{A}$  is partitioned into sub-sections of  $M'$  columns each, where  $M'^L = e^{nR_2}$ .

The  $U$ -codebook consists of all vectors  $\mathbf{A}\beta$  such that  $\beta \in \mathcal{B}_{M,L}$  and the non-zero entries in  $\beta$  are all equal to  $\sqrt{\frac{\sigma^2 + Q}{L}}$ . Let

$$a = \frac{\sigma^2}{\sigma^2 + Q}.$$

**Encoder:** Given source sequence  $\mathbf{X}$ , find the codeword  $\mathbf{U}$  from the SPARC such that  $a\mathbf{U}$  is closest to  $\mathbf{X}$  in Euclidean distance. Specifically, determine

$$\beta^* = \underset{\beta \in \mathcal{B}_{M,L}}{\text{argmin}} \|\mathbf{X} - a\mathbf{A}\beta\|^2.$$

Send the decoder a tuple  $(p_1, \dots, p_L)$  where  $p_i \in \{1, \dots, \frac{M}{M'}\}$  indicates the subsection in the  $i$ th section of  $\mathbf{A}$  where  $\beta^*$  contains a non-zero element. The rate to the decoder is

$$\frac{1}{n} \log \left( \frac{M}{M'} \right)^L = R_1 - R_2.$$

**Decoder:** The decoder first determines the  $n \times M'L$  sub-matrix corresponding to the subsections of  $\mathbf{A}$  indicated by  $p_1, \dots, p_L$ . We denote this sub-matrix  $\mathbf{A}_{bin}$ .  $\mathbf{A}_{bin}$  defines a SPARC  $\mathbf{A}_{bin}\beta$  where  $\beta \in \mathcal{B}_{M',L}$  contains one non-zero value equal to  $\sqrt{(\sigma^2 + Q)/L}$  in each of its  $L$  sections. The decoder now reconstructs  $\hat{\mathbf{U}} = \mathbf{A}_{bin}\hat{\beta}$  where

$$\hat{\beta} = \underset{\beta \in \mathcal{B}_{M',L}}{\text{argmin}} \|\mathbf{Y} - a\mathbf{A}_{bin}\beta\|^2. \quad (14)$$

Finally, the source sequence is reconstructed as

$$\hat{\mathbf{X}} = \left( \frac{1}{Q} + \frac{1}{\sigma^2} + \frac{1}{N} \right)^{-1} \left( \frac{\hat{\mathbf{U}}}{Q} + \frac{\mathbf{Y}}{N} \right) \quad (15)$$

**Error Analysis:** Let  $\delta > 0$  be such that

$$R_1 > \max \left\{ \frac{1}{2} \log \frac{(\sigma^2 + Q)(1 + \delta)}{Q}, 1 - \frac{Q}{(\sigma^2 + Q)(1 + \delta)} \right\} \quad (16)$$

The probability of the error event  $\mathcal{E}$ , can be decomposed as  $P(\mathcal{E}) = P(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3)$  where  $\mathcal{E}_1$  is the event that  $|\mathbf{X}|^2 >$

$\sigma^2(1+\delta)$ ,  $\mathcal{E}_2$  is the event of error at the encoder, and  $\mathcal{E}_3$  the event of error at the decoder. We have

$$P(\mathcal{E}_1) = P(|\mathbf{X}|^2 > \sigma^2(1+\delta)) < \frac{\epsilon}{3} \quad (17)$$

for sufficiently large  $n$  from standard results on large-deviations [36]. Next, we have

$$P(\mathcal{E}_2 | \mathcal{E}_1^c) = P\left(\min_{\beta \in \mathcal{B}_{M,L}} |\mathbf{X} - a\mathbf{A}\beta|^2 > \frac{\sigma^2 Q}{\sigma^2 + Q}\right) < \frac{\epsilon}{3} \quad (18)$$

for sufficiently large  $n$ . This follows from Fact 1 since  $R_1$  and  $b$  satisfy the conditions specified in Fact 1 for compressing source sequences of variance up to  $\sigma^2(1+\delta)$  at distortion-level  $\frac{\sigma^2 Q}{\sigma^2 + Q}$ . Finally, we bound

$$P(\mathcal{E}_2 | \mathcal{E}_1^c, \mathcal{E}_2^c) = P\left(\operatorname{argmin}_{\beta \in \mathcal{B}_{M',L}} \|\mathbf{Y} - a\mathbf{A}_{bin}\beta\|^2 \neq \beta^*\right).$$

Let the number of columns in each sub-section  $M' = L^{b'}$ . Using  $M'^L = e^{nR_2}$  we have

$$b' = \frac{nR_2}{L \log L} = b \frac{R_2}{R_1} > b_0(\text{snr}) \quad (19)$$

where the last inequality is due to the minimum value of  $b$  specified by the theorem. Since

$$R_2 < \frac{1}{2} \log(1 + \text{snr})$$

and  $b' > b_0(\text{snr})$ , the  $n \times M'L$  design matrix  $\mathbf{A}_{bin}$  satisfies the conditions of Fact 2 for signal-to-noise ratio given by (13). Hence for sufficiently large  $n$ ,  $P(\mathcal{E}_2 | \mathcal{E}_1^c, \mathcal{E}_2^c) < \epsilon/3$ . Combining this with (17) and (18), we have  $P(\mathcal{E}) < \epsilon$ . ■

## V. SPARC FOR WRITING ON DIRTY PAPER

The AWGN channel with state is defined by the relation  $Y = X + S + Z$ , where the state  $S \sim \mathcal{N}(0, \sigma_s^2)$  is independent of the additive noise  $Z \sim \mathcal{N}(0, N)$ . There is an average power constraint  $P$  on the input sequence  $\mathbf{X}$ . The state sequence  $\mathbf{S} \sim \text{i.i.d } \mathcal{N}(0, \sigma_s^2)$  is known non-causally at the encoder.

We first review the main ideas behind Costa's capacity-achieving coding scheme [20] for this channel. The state sequence  $\mathbf{S}$  (known only at the encoder) is used in two ways: part of it is used for coding and the rest is treated as noise. Define an auxiliary random variable  $U$  as

$$U = X + \alpha S \quad (20)$$

where  $X \sim (0, P)$  is independent of  $S$  and  $\alpha \in (0, 1)$  is a constant specified later. The channel codebook consists of  $e^{nR_1}$   $U$ -sequences chosen i.i.d  $\mathcal{N}(0, P + \alpha^2 \sigma_s^2)$ . We divide this codebook into  $e^{nR}$  equal-sized bins with each bin representing a message. To transmit message  $m \in \{1, \dots, e^{nR}\}$ , the encoder observes the state sequence  $\mathbf{S}$  and attempts to find a codeword  $\mathbf{U}$  within bin  $m$  whose empirical joint distribution with  $\mathbf{S}$  is close to (20). From rate-distortion theory, this step will be successful if the number of sequences in each bin  $e^{n(R_1-R)}$  is larger than  $e^{nI(U;S)}$ . The encoder then forms the channel input sequence  $\mathbf{X}$  as  $\mathbf{U} - \alpha \mathbf{S}$ .

The channel receives the output sequence  $\mathbf{Y}$  according to

$$Y = X + S + Z = U + (1 - \alpha)S + Z \quad (21)$$

and attempts to decode  $\mathbf{U}$ . This is effectively an AWGN channel decoding operation, which will be successful if  $R_1 < I(U; Y)$ . Combining this with the lower bound  $R_1 - R > I(U; S)$ , we see that any rate  $R < I(U; Y) - I(U; S)$  is achievable. The right-side of the inequality is equal to the channel capacity  $\frac{1}{2} \log(1 + P/N)$  for the joint distribution given by (20) and (21) for  $\alpha = \frac{P}{P+N}$ .

We now show how to implement the above coding scheme with a nested SPARC. Define a nested SPARC with rates  $(R_1, R_1 - R)$  through an  $n \times ML$  matrix  $\mathbf{A}$  with  $M^L = e^{nR_1}$ . As in Section IV, each bin corresponds to a SPARC defined by a sub-matrix of  $\mathbf{A}$ , consisting of  $L$  subsections of  $M'$  columns. We note that  $M'^L = e^{n(R_1-R)}$  which implies that the number of bins is  $(M/M')^L = e^{nR}$ . Thus each message indexes a unique bin of the nested SPARC or equivalently, a unique sub-matrix of  $\mathbf{A}$ .

The relation (20) can be equivalently written in terms of the reverse test channel as

$$S = \kappa U + X' \quad (22)$$

where  $\kappa = \alpha \sigma_s^2 / (P + \alpha^2 \sigma_s^2)$  and  $X' \sim \mathcal{N}(0, \frac{P \sigma_s^2}{P + \alpha^2 \sigma_s^2})$  is independent of  $U$ . Given the message and state sequence  $\mathbf{S}$ , the encoder needs to quantize the state sequence  $\mathbf{S}$  to a codeword  $\kappa \mathbf{U}$  (within the bin indexed by the message) with mean-squared distortion at most  $\text{Var}(X')$ . The SPARC defined by the message bin can reliably perform this quantization if the corresponding sub-matrix of  $\mathbf{A}$  has parameters satisfying the specifications in Fact 1. Using (22), the channel law (21) can be written as

$$Y = U + (1 - \alpha)S + Z = U + (1 - \alpha)\kappa U + (1 - \alpha)X' + Z \quad (23)$$

where  $U, X', Z$  are mutually independent. Thus the decoder has to recover the codeword  $\mathbf{U}$  transmitted over a channel with effective signal-to-noise ratio

$$\text{snr} = \frac{(1 + (1 - \alpha)\kappa)^2 \text{Var}(U)}{(1 - \alpha)^2 \text{Var}(X') + N} = \frac{(1 + (1 - \alpha)\kappa)^2 (P + \alpha^2 \sigma_s^2)}{\frac{(1 - \alpha)^2 P \sigma_s^2}{P + \alpha^2 \sigma_s^2} + N} \quad (24)$$

For all  $R_1 < \frac{1}{2} \log(1 + \text{snr})$ , this step is successful with high probability if the design matrix  $\mathbf{A}$  satisfies the specifications in Fact 2.

The performance of this coding scheme is formalized in the following theorem.

**Theorem 2:** Fix  $\alpha \in (0, 1)$ . Let  $\kappa = \frac{\alpha \sigma_s^2}{P + \alpha^2 \sigma_s^2}$  and  $\text{snr}$  be defined by (24). Fix  $R_1 < \frac{1}{2} \log(1 + \text{snr})$  and  $R_2 > \max\{\frac{1}{2} \log(1 + \frac{\alpha^2 \sigma_s^2}{P}), \frac{\alpha^2 \sigma_s^2}{P + \alpha^2 \sigma_s^2}\}$  such that  $R_1 > R_2$ . There exists a rate  $R_1 - R_2$  code  $\mathcal{C}_n$  with  $P_e(\mathcal{C}_n) < \epsilon$  where  $\mathcal{C}_n$  is defined by a nested sparse regression codebook with rates  $(R_1, R_2)$  whose  $n \times ML$  design matrix satisfies the following:  $M = L^b$  where

$$b > \max\left\{\frac{2.5R_1}{R_2 - \alpha^2 \sigma_s^2 / (P + \alpha^2 \sigma_s^2)}, b_0(\text{snr})\right\}$$

and  $L$  is determined by  $bL \log L = nR_1$ . (The function  $b_0(\cdot)$  is defined in (8)).

*Corollary 2:* Let  $x^* \approx 0.2032$  be the solution of the equation (5). For  $\frac{P\sigma_s^2}{(P+N)^2} \geq \frac{1}{x^*} - 1$ , sparse regression codes achieve the channel capacity  $\frac{1}{2} \log(1 + \frac{P}{N})$ .

*Proof:* It can be verified that when we choose  $\alpha = P/(P+N)$  and  $P, N, \sigma_s^2$  satisfy the above condition, the lower bound on  $R_2$  specified by the theorem becomes  $\frac{1}{2} \log(1 + \frac{\alpha^2 \sigma_s^2}{P})$ . The corollary then follows by choosing  $R_1 = \frac{1}{2} \log(1 + \text{snr}) - \epsilon$  and  $R_2 = \frac{1}{2} \log(1 + \frac{\alpha^2 \sigma_s^2}{P}) + \epsilon$ . This yields an achievable rate  $R_1 - R_2 = \frac{1}{2} \log(1 + P/N) - 2\epsilon$  where  $\epsilon > 0$  can be arbitrarily small. ■

*Proof of Theorem 2:*

Fix block length  $n$  and rates  $R_1, R_2$ . Choose a  $n \times ML$  design matrix  $\mathbf{A}$  with  $M = L^b$  and  $b$  greater than the minimum value specified by the theorem. The  $U$ -codebook consists of all vectors  $\mathbf{A}\beta$  such that  $\beta \in \mathcal{B}_{M,L}$  and the non-zero entries in  $\beta$  are all equal to  $\sqrt{(P + \alpha^2 \sigma_s^2)/L}$ .

We have  $M^L = e^{nR_1}$ , and each section of  $\mathbf{A}$  is partitioned into sub-sections of  $M'$  columns each where  $M'^L = e^{nR_2}$ . Each of the  $e^{n(R_1 - R_2)}$  messages corresponds to a unique tuple  $(p_1, \dots, p_L)$  where  $p_i \in \{1, \dots, \frac{M}{M'}\}$ .

*Encoder:* The message  $(p_1, \dots, p_L)$  indexes an  $n \times M'L$  sub-matrix of  $\mathbf{A}$ . This sub-matrix denoted by  $\mathbf{A}_{bin}$ . Find the codeword  $\mathbf{U}$  from the SPARC  $\mathbf{A}_{bin}$  such that  $\kappa\mathbf{U}$  is closest to  $\mathbf{S}$  in Euclidean distance. Specifically, determine

$$\beta^* = \underset{\beta \in \mathcal{B}_{M',L}}{\operatorname{argmin}} \|\mathbf{S} - \kappa\mathbf{A}_{bin}\beta\|^2$$

and transmit

$$\mathbf{X} = \mathbf{A}_{bin}\beta^* - \alpha\mathbf{S}.$$

*Decoder:* Given channel output  $\mathbf{Y}$ , find the codeword  $\mathbf{U}$  from the SPARC such that  $(1 + (1 - \alpha)\kappa)\mathbf{U}$  is closest to  $\mathbf{Y}$  in Euclidean distance. Specifically, determine

$$\hat{\beta} = \underset{\beta \in \mathcal{B}_{M',L}}{\operatorname{argmin}} \|\mathbf{Y} - (1 + (1 - \alpha)\kappa)\mathbf{A}\beta\|^2.$$

Decode the message as  $(\hat{p}_1, \dots, \hat{p}_L)$  where  $\hat{p}_i \in \{1, \dots, \frac{M}{M'}\}$  indicates the subsection in the  $i$ th section of  $\mathbf{A}$  where  $\hat{\beta}$  contains a non-zero element.

*Error Analysis:* Let  $\delta > 0$  be such that

$$R_2 > \max \left\{ \frac{1}{2} \log \left( (1 + \delta) \frac{(P + \alpha^2 \sigma_s^2)}{P} \right), 1 - \frac{P}{(P + \alpha^2 \sigma_s^2)(1 + \delta)} \right\} \quad (25)$$

The probability of error can be decomposed as  $P(\mathcal{E}) = P(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3)$  where  $\mathcal{E}_1$  is the event that  $|\mathbf{S}|^2 > \sigma_s^2(1 + \delta)$ ,  $\mathcal{E}_2$  is the event of error at the encoder, and  $\mathcal{E}_3$  the event of error at the decoder. We have

$$P(\mathcal{E}_1) = P(|\mathbf{S}|^2 > \sigma_s^2(1 + \delta)) < \frac{\epsilon}{3} \quad (26)$$

for sufficiently large  $n$ . Letting  $M' = L^{b'}$ , we have

$$b' = \frac{nR_2}{L \log L} = b \frac{R_2}{R_1} > \frac{2.5R_2}{R_2 - \alpha^2 \sigma_s^2 / (P + \alpha^2 \sigma_s^2)} \quad (27)$$

where the last inequality is due to the minimum value of  $b$  specified by the theorem. We then have

$$P(\mathcal{E}_2 | \mathcal{E}_1^c) = P \left( \min_{\beta \in \mathcal{B}_{M',L}} |\mathbf{S} - \kappa\mathbf{A}_{bin}\beta|^2 > \frac{P\sigma_s^2}{P + \alpha^2 \sigma_s^2} \right) < \frac{\epsilon}{3} \quad (28)$$

for sufficiently large  $n$ . This follows from Fact 1 since  $R_2$  and  $b'$  satisfy the conditions specified in Fact 1 for compressing sequences  $\mathbf{S}$  of variance up to  $\sigma_s^2(1 + \delta)$  at distortion-level  $\frac{P\sigma_s^2}{P + \alpha^2 \sigma_s^2}$ . Finally,  $P(\mathcal{E}_2 | \mathcal{E}_1^c, \mathcal{E}_2^c)$  is given by

$$P \left( \underset{\beta \in \mathcal{B}_{M,L}}{\operatorname{argmin}} \|\mathbf{Y} - (1 + (1 - \alpha)\kappa)\mathbf{A}\beta\|^2 \neq \beta^* \right) < \epsilon/3 \quad (29)$$

from Fact 2 since  $R_1 < \frac{1}{2} \log(1 + \text{snr})$  and  $b > b_0(\text{snr})$  for  $\text{snr}$  given by (24). Combining (26), (28) and (29), we conclude that  $P(\mathcal{E}) < \epsilon$ . ■

## VI. SPARC FOR GAUSSIAN MULTIUSER CHANNELS

### A. The AWGN Multiple-Access Channel

In a multiple-access channel (MAC), several users simultaneously transmit to a single receiver. For simplicity let us consider the case of two users, each with average power constraint  $P$ , transmitting information at rates  $R_1$  and  $R_2$ , respectively. The receiver of the AWGN MAC observes the output

$$Y = X_1 + X_2 + Z$$

where  $X_1, X_2$  denote the channel inputs of the two transmitters and  $Z \sim \mathcal{N}(0, N)$  is the channel noise independent of  $X_1$  and  $X_2$ . The capacity region for this channel is well-known [35] and is given by

$$\begin{aligned} R_1 &< \frac{1}{2} \log(1 + \frac{P}{N}), \quad R_2 < \frac{1}{2} \log(1 + \frac{P}{N}), \\ R_1 + R_2 &< \frac{1}{2} \log(1 + \frac{2P}{N}). \end{aligned} \quad (30)$$

We now show how to achieve the corner points of this rate region using SPARCs. The remaining rate points in the region can be achieved through time-sharing. The key observation is that the corner points can be achieved using a pair of *point-to-point* channel codes [37].

Consider a rate pair

$$R_1 < \frac{1}{2} \log(1 + \frac{P}{P + N}), \quad R_2 < \frac{1}{2} \log(1 + \frac{P}{N}).$$

Fix codeword length  $n$  and choose a rate  $R_1$  SPARC for transmitter 1 using an  $n \times M_1 L_1$  design matrix  $\mathbf{A}_1$ , with  $M_1^{L_1} = 2^{nR_1}$  and  $\mathbf{A}_1$  satisfying the specifications of Fact 2 for signal-to-noise ratio  $\frac{P}{P + N}$ . Similarly, chose a rate  $R_2$  SPARC for transmitter 2 using an  $n \times M_2 L_2$  design matrix  $\mathbf{A}_2$ , with  $M_2^{L_2} = 2^{nR_2}$  and  $\mathbf{A}_2$  satisfying the specifications of Fact 2 for signal-to-noise ratio  $\frac{P}{N}$ . The codewords  $X_1 = \mathbf{A}_1 \beta_1$  and  $X_2 = \mathbf{A}_2 \beta_2$  chosen by the respective users, are transmitted through the channel. The non-zero values in both  $\beta_1$  and  $\beta_2$  are set to  $\sqrt{P/L}$ . The receiver obtains

$$\mathbf{Y} = \mathbf{A}_1 \beta_1 + \mathbf{A}_2 \beta_2 + \mathbf{Z}$$

and uses a successive cancellation strategy. It first decodes the message of transmitter 1 as  $\hat{\beta}_1$ , effectively treating  $\mathbf{A}_2\beta_2 + \mathbf{Z}$  as noise. This step will be successful with high probability since the signal-to-noise ratio is  $P/(P+N)$  and  $R_1 < \frac{1}{2}\log(1 + \frac{P}{P+N})$ . In the second step, the receiver decodes  $\beta_2$  from the residue  $\mathbf{Y} - \mathbf{A}_1\hat{\beta}_1$  which equals  $\mathbf{A}_2\beta_2 + \mathbf{Z}$  if the first step was successful, i.e.,  $\hat{\beta}_1 = \beta_1$ . The second step will be successful with high probability since  $R_2 < \frac{1}{2}\log(1 + \frac{P}{N})$ . The other corner point of the rate region can be achieved by exchanging the roles of  $X_1$  and  $X_2$ .

### B. The AWGN Broadcast Channel

Consider the two-receiver scalar Gaussian broadcast channel where the outputs of the two receivers are related to the channel input  $X$  as

$$Y_1 = X + Z_1, \quad Y_2 = X + Z_2.$$

$X$  has average power constraint  $P$  and the channel noises  $Z_1, Z_2$  are independent zero mean Gaussian random variables with variances  $N_1$  and  $N_2$ , respectively. Without loss of generality, we assume that  $N_2 \geq N_1$ . The capacity region for this channel is given by [35]

$$R_1 < \frac{1}{2} \log \left( 1 + \frac{\alpha P}{N_1} \right), \quad R_2 < \frac{1}{2} \log \left( 1 + \frac{(1-\alpha)P}{\alpha P + N_2} \right). \quad (31)$$

This capacity region is achievable through a SPARC coding scheme similar to the one for the AWGN MAC. Consider a rate pair  $(R_1, R_2)$  satisfying (31). Fix codeword length  $n$  and choose a rate  $R_1$  SPARC for transmitter 1 using an  $n \times M_1 L_1$  design matrix  $\mathbf{A}_1$ , with  $M_1^{L_1} = 2^{nR_1}$  and  $\mathbf{A}_1$  satisfying the specifications of Fact 2 for signal-to-noise ratio  $\frac{\alpha P}{N_1}$ . Similarly, chose a rate  $R_2$  SPARC for transmitter 2 using an  $n \times M_2 L_2$  design matrix  $\mathbf{A}_2$ , with  $M_2^{L_2} = 2^{nR_2}$  and  $\mathbf{A}_2$  satisfying the specifications of Fact 2 for signal-to-noise ratio  $\frac{(1-\alpha)P}{\alpha P + N_2}$ . The input sequence is generated as  $\mathbf{X} = \mathbf{A}_1\beta_1 + \mathbf{A}_2\beta_2$  where  $\beta_1, \beta_2$  represent the messages of the two users. The non-zero values in  $\beta_1$  and  $\beta_2$  are set to  $\sqrt{\alpha P/L}$  and  $\sqrt{(1-\alpha)P/L}$ , respectively. The receivers obtain

$$\mathbf{Y}_1 = \mathbf{A}_1\beta_1 + \mathbf{A}_2\beta_2 + \mathbf{Z}_1, \quad \mathbf{Y}_2 = \mathbf{A}_1\beta_1 + \mathbf{A}_2\beta_2 + \mathbf{Z}_2.$$

Receiver 2 decodes  $\beta_2$  treating  $\mathbf{A}_1\beta_1 + \mathbf{Z}_2$  as noise. This will be successful with high probability since  $R_2 < \frac{1}{2} \log_2 \left( 1 + \frac{(1-\alpha)P}{\alpha P + N_2} \right)$ . Receiver 1 can also decode  $\beta_2$  with high probability since its since  $N_1 \leq N_2$ . Hence the residue  $\mathbf{Y}_1 - \mathbf{A}_2\hat{\beta}_2$  at receiver 1 will be equal to  $\mathbf{A}_1\beta_1 + \mathbf{Z}_1$  with high probability. Receiver 1 can then reliably decode  $\beta_1$  from this residue since the rate  $R_1 < \frac{1}{2} \log(1 + \frac{\alpha P}{N_1})$ .

### VII. CONCLUSION

The results of [14]–[19] showed that the sparse regression framework can be used to design rate-optimal codes for point-to-point source and channel coding with computationally efficient encoders and decoders. In this paper, we showed how these source and channel codes can be combined to implement random binning and superposition. These two

techniques have been fundamental ingredients of rate-optimal coding schemes for a wide range of problems in network information theory. The next goal is a precise performance analysis of the computationally feasible versions of the coding schemes presented here. Constructing a library of efficient sparse regression modules to perform source coding, channel coding, binning and superposition will pave the way for fast, rate-optimal codes for several network problems such as multiple description coding, lossy distributed source coding, interference channels and relay channels.

### REFERENCES

- [1] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. 18, pp. 2 – 14, Jan 1972.
- [2] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, pp. 471 – 480, Jul 1973.
- [3] A. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 21, pp. 294 – 300, May 1975.
- [4] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, pp. 1 – 10, January 1976.
- [5] S. Gelfand and M. Pinsker, "Coding for a channel with random parameters," *Problems of Control and Information*, vol. 9, pp. 19 – 31, January 1980.
- [6] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3051 – 3073, July 2009.
- [7] S. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inf. Theory*, vol. 56, pp. 1751 – 1768, April 2010.
- [8] S. Korada and R. Urbanke, "Polar codes for slepian-wolf, wyner-ziv, and gelfand-pinsker," in *IEEE Information Theory Workshop*, Jan. 2010.
- [9] E. Abbe and E. Telatar, "Polar codes for the m-user MAC and matroids," *IEEE Trans. Inf. Theory*, 2011.
- [10] S. Kudekar, T. Richardson, and R. L. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Int. Symp. Inf. Thy*, 2012. Also on <http://arxiv.org/abs/1201.2999>.
- [11] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1250 – 1276, June 2002.
- [12] M. Eyuboglu and J. Forney, G.D., "Lattice and trellis quantization with lattice- and trellis-bounded codebooks-high-rate theory for memoryless sources," *IEEE Trans. Inf. Theory*, vol. 39, pp. 46 – 59, Jan 1993.
- [13] U. Erez and R. Zamir, "Achieving  $1/2 \log(1 + \text{snr})$  on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293 – 2314, 2004.
- [14] A. Barron and A. Joseph, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Transactions on Information Theory*, vol. 58, pp. 2541 – 2557, Feb 2012.
- [15] A. Barron and A. Joseph, "Toward fast reliable communication at rates near capacity with Gaussian noise," in *IEEE Int. Symp. Inf. Theory*, 2010.
- [16] A. Joseph and A. Barron, "Fast sparse superposition codes have exponentially small error probability for  $R < C$ ," *Submitted to IEEE Trans. Inf. Theory*, 2012. <http://arxiv.org/abs/1207.2406>.
- [17] I. Kontoyiannis, K. Rad, and S. Gitzenis, "Sparse superposition codes for Gaussian vector quantization," in *2010 IEEE Inf. Theory Workshop*, p. 1, Jan. 2010.
- [18] R. Venkataramanan, A. Joseph, and S. Tatikonda, "Gaussian rate-distortion via sparse linear regression over compact dictionaries," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2012. Also on <http://arxiv.org/abs/1202.0840>.
- [19] R. Venkataramanan, T. Sarkar, and S. Tatikonda, "Lossy compression via sparse linear regression: Computationally efficient encoding and decoding," <http://arxiv.org/abs/1212.1707>.
- [20] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, pp. 439 – 441, May 1983.
- [21] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," *IEEE Trans. Inf. Theory*, vol. 49, pp. 626 – 643, Mar 2003.



- [22] D. Rebollo-Monedero, R. Zhang, and B. Girod, "Design of optimal quantizers for distributed source coding," in *Data Compression Conference*, pp. 13 – 22, March 2003.
- [23] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes," *IEEE Trans. Communications*, vol. 57, pp. 376 – 387, Feb. 2009.
- [24] M. Wainwright and E. Martinian, "Low-density graph codes that are optimal for binning and coding with side information," *IEEE Trans. Inf. Theory*, vol. 55, pp. 1061 – 1079, March 2009.
- [25] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3401–3416, 2005.
- [26] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3820 – 3833, Nov. 2005.
- [27] Q. Spencer, C. Peel, A. Swindlehurst, and M. Haardt, "An introduction to the multi-user mimo downlink," *IEEE Communications Magazine*, vol. 42, pp. 60 – 67, Oct. 2004.
- [28] P. Moulin and R. Koetter, "Data-hiding codes," *Proceedings of the IEEE*, vol. 93, pp. 2083 – 2126, Dec. 2005.
- [29] B. Chen and G. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, pp. 1423 – 1443, May 2001.
- [30] J. O'Sullivan, P. Moulin, and J. Ettinger, "Information theoretic analysis of steganography," in *IEEE Int. Symp. on Inf. Theory*, 1998.
- [31] S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1181 – 1203, May 2003.
- [32] R. Barron, B. Chen, and G. Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1159 – 1180, May 2003.
- [33] Y. Sun, Y. Yang, A. Liveris, V. Stankovic, and Z. Xiong, "Near-capacity dirty-paper code design: A source-channel coding approach," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3013 – 3031, July 2009.
- [34] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3417–3432, 2005.
- [35] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, Inc., 2001.
- [36] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, 1998.
- [37] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, pp. 364 – 375, Mar 1996.